行政院國家科學委員會專題研究計畫　期中進度報告

應用可逆躍式馬可夫鏈蒙第卡羅法於潛在變數模型的參數推估(1/2)

計畫類別：個別型計畫
計畫編號：NSC94-2118-M-009-008-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學統計學研究所

計畫主持人：黃冠華

計畫參與人員：陳穗碧、劉育倫、鄭俊凱、謝志強、謝國偉

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 95 年 5 月 11 日

目　　錄

一、 中英文摘要

中文摘要

在許多醫學研究中，我們常因為花費或時間等的問題而觀察不到最想要的觀察值。此時，一個有效的替代變數便常被用來取代這個我們所觀察不到的結果。潛在變數模型被認為是一個分析替代變數最有效的統計方法。我的研究主要是針對當我們用一系列的問卷問題作為替代變數，且假設那個觀察不到的潛在變數是一離散形的隨機變數其維度為 J。

在過去，研究者常應用下列方法來做潛在變數模型的參數估計：首先根據適切度檢定或 AIC、BIC 判定準據來選取潛在變數的維度，然後再假設維度為已知並做參數估計。此一“兩階段”的方法雖然方便，但並不有效率且常會誤導結論。本計劃將嘗試同時估計潛在變數的維度和模型參數。傳統頻率論者的方法並無法做這種共同分析，因此我們將採用貝氏分析法。Green (1995) 所提出的可逆躍式馬可夫鏈蒙第卡羅法將被用來做此一貝氏共同分析。

關鍵詞：貝氏分析、潛在變數模型、可逆躍式馬可夫鏈蒙第卡羅法、替代變數。

Abstract

Many concepts in medical research are unobservable, hence valid surrogates must be measured in place of these concepts. Models that permit exploration of relationships between unobservable concepts and their surrogates are referred to as latent class models. My research is focused on analyzing data collected in situations where investigators use multiple discrete indicators as surrogates, for example, a set of questionnaires and assume an underlying categorical latent variable with, say, J “classes”.

Most previous latent class estimation considers models for different numbers of classes separately and use significance tests or information criteria (such as AIC or BIC) to infer the number of classes. This two-stage approach is inefficient and may create misleading results. Joint inferences on the number of classes and model parameters are preferable on the ground of convenience, accuracy and flexibility. Traditional frequentist likelihood-based approaches do not allow this joint analysis, but recent advances in Bayesian inferences provide possible solutions. Green (1995) proposed the reversible jump Markov chain Monte Carlo (RJMCMC) method, which offers a general framework for construction of reversible Markov chain samplers that jump between parameter subspaces of different dimensionality. In this project, we propose to implement the RJMCMC method to perform the joint estimation of the number of classes and model parameters.

Keywords: Bayesian analysis, latent variable model, reversible jump Markov chain Monte Carlo, surrogate endpoint.

二、 緣由與目的

Latent class analysis (LCA), originally described by Green (1951) and systematically developed by Lazarsfeld and Henry (1968), Goodman (1974), has been found useful for classifying subjects based on their responses to a set of categorical items. This project studies a more general latent class model proposed by Huang and Bandeen-Roche (2004), which incorporates covariate effects both on the latent variable and the measured items themselves.

To reduce complexity and enhance interpretability, one usually fixes the number of levels or "classes" in a given latent class model and does the parameter estimation under the fixed number of classes. Standard practice is to base selection on either the Pearson χ^2 or the likelihood ratio goodness of fit test, and to fix J at the lowest number of classes that yields acceptable fit (Goodman 1974, Formann 1992). Instead of testing the goodness of fit of a specified model, we might use a criterion for selecting among different numbers of classes. The AIC (Akaike 1987) and BIC (Schwarz 1978) criterion, which trades off the value of the likelihood at the maximum likelihood solution and the number of estimated parameters, are commonly used approaches.

Above two-stage approaches are inefficient and may create misleading results. Joint inferences on the number of classes and model parameters are preferable because it is convenient, accurate and flexible. Traditional frequentist likelihood-based approaches do not allow this joint analysis, but recent advances in Bayesian inferences provide possible solutions. Green (1995) proposed the reversible jump Markov chain Monte Carlo (RJMCMC) method, which offers a general framework for construction of reversible Markov chain samplers that jump between parameter subspaces of different dimensionality. Latent class models with different numbers of classes correspond to parameter subspaces of different dimensionality. RJMCMC thus provides a solution to jointly estimate the number of classes and model parameters.

Richardson and Green (1997) made use of RJMCMC to model the number of components and the mixture component parameters jointly for a finite mixture with univariate normal mixtures. Latent class models can be viewed as a finite mixture with multinomial mixtures. As we know, none of published MCMC applications has deal with this situation. In this project, we propose to implement the RJMCMC method to perform the joint estimation of the number of classes and model parameters of the latent class regression model proposed by Huang and Bandeen-Roche (2004).

三、 結果與討論

Up to now we have done the followings:

1) Have built up the initial Bayesian models and RJMCMC algorithm for latent class regressions.
2) Have implemented the proposed algorithm through R, C and BUGS, and have performed the simulation studies.
3) Modifying the original algorithm based on the results of simulation studies.

We are currently writing up a manuscript for initial results:

Title: Bayesian inferences on latent class regression with an unknown number of components via reversible jump Markov chain Monte Carlo

Abstract:
Latent class models have proven useful for exploring the relationships between unobservable concepts and their surrogates. This paper is concerning a general latent class model proposed by Huang and Bandeen-Roche (2004), which incorporates covariate effects both on the unobservable concept and the measured items themselves. Most previous latent class estimation considers models for different numbers of classes separately and use significance tests or information criteria (such as AIC or BIC) to infer the number of classes. This two-stage approach is inefficient and may create misleading results. Joint inferences on the number of classes and model parameters are preferable on the ground of convenience, accuracy and flexibility. In this paper, we implement the reversible jump Markov chain Monte Carlo (RJMCMC) method (Green 1995; Richardson and Green 1997) to perform the joint estimation of the number of classes and model parameters. Bayesian model setting and RJMCMC algorithm for latent class regressions are proposed. Simulation studies show reasonable accuracy on parameter estimation.

四、 計畫成果自評

We have met all proposed aims set in the proposal. Further things that need to be done are:

1) Will finalize the manuscript and submit it to the SCI journal.
2) Will further modify the proposed model and establish a most efficient algorithm.
3) Will perform the sensitivity analysis to evaluate how sensitive the parameter estimates are to specified priors, and select appropriate priors accordingly.
4) The R package for performing the proposed algorithm will be ready for download.

五、 參考文獻

Akaike H (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332.

Formann AK (1992). Linear logistic latent class analysis for polytomous data. *Journal of the*

American Statistical Association, 87, 476-486.

Goodman LA (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

Green BF (1951). A general solution of the latent class model of latent structure analysis and latent profile analysis. *Psychometrika*, 16, 151-166.

Green PJ (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.

Huang GH, Bandeen-Roche K (2004). Building an identifiable latent variable model with covariate effects on underlying and measured variables. *Psychometrika*, 69, 5-32.

Lazarsfeld PF, Henry NW (1968). *Latent Structure Analysis*. New York: Houghton-Mifflin.

Richardson S, Green PJ (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59, 731-792.

Schwarz G (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.